PRELIMINARY WRITTEN SUMMARY
REVIEW OF MODELING FRAMEWORK DESIGN AND QUALITY ASSURANCE PROJECT PLAN FOR MODELING STUDY OF PCB CONTAMINATION IN THE HOUSATONIC RIVER

Prepared by:
Peter Shanahan, Ph.D., P.E.
HydroAnalysis, Inc.
481 Great Road, Suite 3
Acton, Massachusetts 01720
(978) 263-1092
shanahan@ma.ultranet.com

INTRODUCTION

This review provides my preliminary thoughts concerning the proposed modeling study of contamination of the Housatonic River by polychlorinated biphenyls (PCBs). It is based on a review of the Modeling Framework Design (MFD) (Beach *et al.*, 2000a), its accompanying Quality Assurance Project Plan (QAPP) (Beach *et al.*, 2000b), and various supporting documentation and data provided by the U.S. Environmental Protection Agency (EPA). The review is preliminary in that I continue to review materials provided by the EPA, including the recently received "EPA Response to Peer Review Panelist Questions on the Housatonic River Modeling Framework Design." Nonetheless, I have tried to convey several "big-picture" issues that will be the most critical aspect of my final review.

MODEL RELIABILITY

I have grave concerns that the models that will result from this effort may not be sufficiently reliable to form the basis for environmental cleanup decisions that have major implications for public and environmental health and could involve expenditures in the range of several hundred million dollars. The reasons for my concerns can be categorized under the following general issues: excessive model complexity, inadequate calibration, and uncertain predictive ability. Each is discussed in turn in the following.

MODEL COMPLEXITY

Despite lip service to model parsimony in Section 3.1.1 of the MFD, the models proposed for this study appear to have been developed with the philosophy that the more detailed and complex the formulation, the better. I see intrinsic difficulties in this approach, although my aversion is more visceral than analytical. Fortunately, Bruce Beck has provided an analysis of the issue that is both philosophical and rigorous in two recent papers (Beck, 1987; Beck and Halfon, 1991). The problems he identifies can be generally described as an inability to develop a proper model structure, uncertainty in the model parameters, and poor predictive capability.

The pertinence of the issue of proper model structure is demonstrated in a recent review of the models developed to assess PCB contamination in the Fox River in Wisconsin (Tracy and Keane, 2000). A physically-based (as opposed to empirical) model of the Fox River was developed by the Wisconsin Department of Natural Resources (WDNR), calibrated to field data for PCBs and suspended solids, and apparently used as the basis for a draft Superfund feasibility study (FS) of alternative strategies to remediate the river. Alternative cleanup strategies range from no action to dredging and contained disposal of contaminated sediments at an estimated cost of $720 million (WDNR, 1999). Subsequent to the publication of the models and draft FS, the WDNR model has been examined in detail by consultants to the industries identified as Potentially Responsible Parties (PRPs). The consultants identified a serious structural flaw in the model. Numerical dispersion arising from the representation of PCB transport in the sediments produces a large artificial flux of PCBs from deep sediment layers to the sediment surface in the WDNR model. As the result of this and other differences in the models, the WDNR model predicts that 70% more PCBs will be discharged from the Fox River than does an alternative model developed by the PRPs' consultants. WDNR is in the process of correcting its model, but has not yet reported the effect on the model predictions. There appears to be no consensus among the various parties on a best model or best modeling approach.

The specific flaw identified in the WDNR Fox River model is known to the modelers working on the Housatonic River and will not be repeated. Moreover, the Housatonic River modelers also know the recommendations of the Fox River Peer Review Panel, although the study is cited only in the response to questions from this Peer Review Panel and not in the MFD or QAPP. Regardless of the specific flaw in the WDNR model and its correction, the Fox River example reveals a fundamental failing in the application of these highly complex models: the models are over-parameterized such that even an intrinsically flawed model can be "calibrated" to field data. I suspect that with the number of parameters included in the AQUATOX model, it could be satisfactorily calibrated to any time-series data set, including the Dow Jones Industrial Average.

Beck's review of model uncertainty leaves me pessimistic—he states for example that "Overparameterization seems both intrinsic and an intractable problem" (Beck and Halfon, 1991). He also makes clear that a "physics-based mechanistic" approach is hardly a panacea (Beck, 1987). It is noteworthy that he cites the predecessor model of AQUATOX as an example of the misguided physics-based approach. In essence, Beck argues that these models cannot be tested by the scientific method because they cannot be shown to be false. There are too few field data to disprove the many subordinate hypotheses and parameters embedded in these complex models. Thus, one cannot test hypotheses as is incumbent with the scientific method.

Despite this overall pessimism, Beck (1987) does provide some optimistic observations. In particular, he distinguishes simulation of hydrology (and by implication, hydrodynamics) from simulation of water quality. He cites the longer and more intensive study of hydrologic processes compared to the more ephemeral attention to water quality (which has shifted attention over the years from BOD/DO, to eutrophication, to acid precipitation, and now to toxic chemicals). As a result of more intensive historical examination, the hydrological and hydrodynamic processes are better understood, more certainly parameterized, and better identified.

Another seemingly pessimistic aspect of Beck's analysis is the fact that his focus is primarily on eutrophication modeling, which is only a subset of the modeling exercise proposed here and thus less complex. However, as Beck (1987, Figure 15) illustrates, the eutrophication problem has some intrinsic difficulties that may not be shared by the PCB problem. Specifically, the prediction of eutrophication involves the translation of relatively steady meteorological and nutrient loading forcing functions into episodic algal blooms—an abrupt transient response that bears little resemblance to the character of the forcing functions. The PCB problem can be idealized as a relatively better behaved problem: namely the exponential depletion and burial of mass over decadal time scales. Indeed, with this conceptual model, I wonder if a calibrated exponential decay coefficient could be as reliable a predictor as the modeling effort proposed here. The flaw in the exponential conceptual model is, of course, the potentially great influence of unusually high flow events. However, the predictive ability of the proposed modeling framework for high flow events is perhaps the single greatest uncertainty in the model.

Another aspect of model construction concerns me on a more practical level: the project team's failure to consider important data available from Massachusetts sources.

The following potentially valuable data were not considered and apparently were unknown to the modeling team:

Data Source	Data type
Massachusetts Department of Environmental Protection	Past water-quality assessments in 1997-1998, 1992, 1985, 1976-1978, 1974, 1968-1969 that variously included water-quality sampling, wastewater discharge surveys, biota sampling, sediment sampling, and probably time-of-travel and other hydrodynamic field studies.
MassGIS	Geographic information system coverages of soil types, land use, wetlands, surficial geology, topography, aerial photography, and other geographical features.
Federal Emergency Management Agency	Hydrologic and hydraulic studies conducted under the Flood Insurance Program, possibly high-water mark surveys after flood events

These data are likely to assist in the formulation of food-web relations, construction of historical simulations, calibration of the hydrodynamic and hydrologic models, and calibration of the phytoplankton component of the water-quality model. I am concerned that the failure to search for and incorporate these important past data could betray a false confidence in the proposed modeling framework: in other words, that the modelers could have concluded that their models are so good and so fundamentally sound, that they do not need to exert every effort to locate the best data.

INADEQUATE CALIBRATION

Beck's analysis makes clear that a "good" calibration in and of itself does not guarantee that the model have been defined correctly or that its parameter values are reasonable. Moreover, simple statistical measures may be misleading as was demonstrated by the apparently successful calibration of the incorrect Fox River model.

A partial remedy to the calibration issue is achieved in the Housatonic River MFD and QAPP by disaggregating certain process and calibrating those separately. For example, completion of an independent sediment erosion study (Gailani *et al.*, 2000) provides some assurance that even if the entire modeling framework cannot be calibrated and verified, important sediment processes can be adequately represented. The more such "subcalibrations" can be completed, the more confidence can be ascribed to the overall model. This approach seems to be endorsed by the Fox River Peer Review Panel as well (Tracy and Keane, 2000). Thus, I would encourage close scrutiny of the models for subprocesses that can be isolated and tested.

The following discusses the calibration for the individual model components: HSPF, EFDC, AQUATOX, and their linkage.

Given its long history of use, and the fact that it has been tested and found successful in at least one post-audit (Hartigan, 1983), I have more confidence in the use of HSPF than of the other models. Nonetheless, I found the description of the HSPF calibration in Section 4.5 of the QAPP to be inadequate. The text is a litany of the parameters that will be adjusted, and the sequence in which they will be adjusted, but provides little information on the field data to which the data will be calibrated. It is obvious however that the available data will allow only gross calibration of the model based primarily on mainstem hydrologic stations. Data collected in 1999 are apparently the only tributary data available, although the description of the datasets is sparse. The Supplemental Investigation Workplan (Weston, 2000, Figure 5.3-1) shows a limited number of tributary stations, however. This data set will be used for calibration, leaving no tributary dataset available for validation. To the extent that PCB contributions from the watershed are identified as an important contributor to the overall PCB balance, supplementary tributary datasets that include high-flow events should be collected for model validation.

The EFDC model seems to be so computationally burdensome that the ability to calibrate the model is constrained. I am, for example, concerned that restrictions in the spatial extent of the EFDC model simultaneously restrict the completeness of its calibration. I was struck by the fact that sediment data were collected on a daily basis at the U.S. Geological Survey gaging station on the Housatonic River in Great Barrington from April 1979 to September 1980. (The flow record extends from 1913 to the present.) Unfortunately, the period of record for sediment data does not include flow events greater than about a three-year event. Nonetheless, these daily data provide a potentially valuable calibration dataset that will not be utilized because the EFDC model does not extend far enough downstream. I think consideration should be given to extending the model to Great Barrington to enable use of the Great Barrington flow and sediment records. Extending the model would also make the HSPF and EFDC models spatially consistent and provide the opportunity for cross-comparison of the model results.

The dataset for calibration of the EFDC model is insufficient to test a critical aspect of the model: its ability to accurately predict the effect of high-flow events. The potential impact of future high flow events is a matter of speculation at other riverine PCB contamination sites, but has nonetheless been held out as a rationale for expensive dredging remediation alternatives. Until a sufficiently extreme flow event presents itself and is thoroughly monitored, the predictive abilities of the Housatonic River model will remain in doubt.

The AQUATOX model is a classic example of an overparameterized and ill-defined model. The QAPP lists numerous statistical and qualitative measures that will be used to assess the calibration of the AQUATOX model, but these will provide no assurance that the model calibration is unique or correct. I see no escape from this dilemma: rather, the model predictions must be qualified by explicit calculation of the model uncertainty so that decision makers can judge the confidence with which the model results can be used.

The final aspect of calibration could also be considered a model structure question: the linkage of the various model components. None of the models are consistent in their representation of the pertinent state variables, and translation is required to transfer information from HSPF to EFDC and from HSPF and EFDC to AQUATOX. These transfers will rely, at least in part, upon approximate conversion formulae described in the QAPP. There is a concern here that the already imperfect predictions from HSPF and EFDC could be further compromised by imperfect conversions. To the extent possible, independent evaluation and calibration of the conversion algorithms should be accomplished.

UNCERTAIN PREDICTIVE ABILITY

The proposed model will be used to assess both historical and future conditions. Historical issues included in MFD objectives entail quantifying the historical and relative contributions of various sources of PCBs on ambient water quality, bed sediment, and bioaccumulation. The intended duration and character of these historical evaluations is not detailed, but it would nonetheless seem that the database as described in the MFD and QAAP is insufficient to support these. Specifically, suspended solids (particularly organic particulates) could have been significantly influenced in the past by municipal and industrial wastewater discharges. Papermill effluents, which affect the river downstream of Woods Pond, are likely to have been particularly significant sources of organic suspended solids prior to the implementation of the Clean Water Act. The database as described in the documents excludes the historical information from Massachusetts DEP needed to characterize historical river conditions. There may as well be historical information that predates Massachusetts DEP—a good source of such information is the Massachusetts State House Library in Boston. In any case, calibration to today's post-Clean-Water-Act conditions does not imply a model capable of accurately simulating pre-Clean-Water-Act conditions.

Prognostic simulations included within the model objectives include estimates of the effects of selected remedial actions, of natural recovery, and of extreme storm events. Remedial alternatives in the case of river sediments necessarily include dredging. Since opinions on dredging run the gamut from a remedial panacea that is practically pre-ordained to counterproductive folly that only spreads and worsens the contamination, it is important to foresee as a part of the modeling framework design how dredging will be evaluated with the model. The representation of dredging, though a critical factor in the eventual model-based decision-making, is absent from the MFD. I am pessimistic that a fair assessment of the effects of dredging will be accomplished without its prior explicit specification. How will sediment disruption and collateral release during dredging be modeled? Is this modeling framework adequate to address that question? I fear that if these questions are not answered upfront, the eventual model application will founder on disagreements over its application.

There is also the considerable uncertainty, alluded to earlier, in the use of the models to forecast behavior of the system at two extremes: the high-flow event, and the system response over decades. The short record available for calibration, and the absence of high-flow events from that record, are ominous for both of these capabilities. By definition, it is unlikely that a high-flow event will be observed during the data-collection period, and it is impractical to extend the observation period to decades. Thus, neither of these critical needs is likely to be addressed. This is not just an issue for model calibration but also of model identification. Specifically, there is considerable uncertainty that the EFDC model can adequately represent the transition from in-bank flow within the meandering stream channel to overbank flow through the floodplain.

How then, in the absence of adequate data, can the uncertainty in these extreme predictions be factored into decision-making? It seems that there is only one possibility, namely to include within the modeling exercise explicit tasks to estimate and report the uncertainty bounds on the model estimates. This task is given only token attention in the MFD, and even then indicated as an optional task that "may" be done. In the absence of explicit evaluation of uncertainty, the inevitable result is that described by Beck and Halfon (1991):

> There must have been many occasions on which a large model, confronted with hopelessly inadequate data, has been accepted as an appropriate tool for

prediction, and without any attempt at quantification of the error attaching to the predictions so obtained.

In effect, without explicit evaluation of uncertainty, the model is implicitly portrayed as certain. For example, the tracing of a single solid predicted-results line on a graph, without error bounds, would represent an implicitly certain forecast. The risk is that unnecessary exposure to PCBs and unwarranted expenditures would result from the selection of a predicted outcome that does not in actual fact differ from an unselected alternative.

The MFD indicates that the complexity of the models, and particularly the EFDC code, precludes rigorous evaluation of uncertainty. In its place, sensitivity tests are proposed. I found the description of the approach to be inadequate however. It seems to me that it is at least as important to communicate to decision makers the level of uncertainty in the model predictions, as it is to communicate the predictions. Thus, I believe more explicit and formalized procedures for the calculation and communication of uncertainty should be incorporated into the MFD.

Some of the other members of this panel have asked about a contingency plan in the event the models are found inadequate to meet the study objectives. While I do not share concerns about this issue *per se*, I have a closely related question, namely: With what confidence can the models be used to answer the questions posed by the study objectives? In essence, I assume the modelers can get *an* answer—my concern is the utility and reliability of that answer.

MISCELLANEOUS ITEMS

The following is a list of miscellaneous items that I flagged for attention during review of the documents and have not previously communicated.

- The QAAP does not appear to include procedures to specifically check model input data. Section 11 appears to touch on this, but should also specify that all input data time series be plotted for visual inspection and cross-checking.

- The response to Peer Review Panel questions regarding past model applications is primarily a list of past applications and calibrations. It would be useful to indicate which, if any, of these are post-audits of these models. Post-audits seem to me to provide a more convincing test of the model than a calibration/validation exercise.

- My previously submitted question 35 (PS17) was not specific enough. My original question was: How will dispersion coefficients be determined? My question should have been: How will dispersion coefficients for use in AQUATOX box model be determined? Specifically, presuming the hydrodynamic information from EFDC will be used to derive dispersion coefficients, will account be taken of the numerical dispersion introduced by the AQUATOX spatial discretization when translating the EFDC dispersion coefficients for use in AQUATOX? How big will the AQUATOX spatial segments be? How much numerical dispersion will the segment size introduce?

- What is the spatial distribution of the tributary datasets that will be used for calibration/testing of HSPF? Do these datasets provide an adequate cross section of the basin in terms of topography, geology, soils, and land use?

- My previously submitted question 84 (PS19) was in error as to the timing of the FEMA survey. I have since talked to participants in the survey and it is the storm of January 1979 that was followed by a high-water mark survey in at least some Massachusetts rivers. Nonetheless, FEMA may still be a useful source of information on past Housatonic River floods.

- Regarding historical bathymetric data for Woods Pond, the Massachusetts Division of Fisheries and Wildlife has in the past prepared contour maps for most Massachusetts ponds and lakes. It may be that they have surveyed Woods Pond.

REFERENCES

Beach, R.B., P.M. Craig, R. DiNitto, A.S. Donigian, G. Lawrence, R.A. McGrath, R.A. Park, A. Stoddard, S.C. Svirsky, W.D. Tate, and C.M. Wallen, 2000a. Modeling Framework Design, Modeling Study of PCB Contamination in the Housatonic River. U.S. Army Corps of Engineers, New England District, Concord, Massachusetts and U.S. Environmental Protection Agency, Region I, Boston, Massachusetts. October 2000.

Beach, R.B., J.S. Clough, P.M. Craig, A.S. Donigian, R.A. McGrath, R.A. Park, A. Stoddard, S.C. Svirsky, and C.M. Wallen, 2000b. Quality Assurance Project Plan, Modeling Study of PCB Contamination in the Housatonic River. U.S. Army Corps of Engineers, New England District, Concord, Massachusetts and U.S. Environmental Protection Agency, Region I, Boston, Massachusetts. October 2000.

Beck, M.B., 1987. Water Quality Modeling: A Review of the Analysis of Uncertainty. *Water Resources Research*, Vol. 23, No. 8, Pp. 1393-1442. August 1987.

Beck, M.B., and E. Halfon, 1991. Uncertainty, Identifiability and the Propagation of Prediction Errors: A Case Study of Lake Ontario. *Journal of Forecasting*, Vol. 10, No. 1 & 2, Pp. 135-161.

Gailani, J.Z., S.J. Smith, M.G. Channell, G.E. Banks, D.B. Brister, R.A. Jepsen, and J.D. Roberts, 2000. Sediment Erosion Study for the Housatonic River, Massachusetts

(Draft). U.S. Army Corps of Engineers, New England District, Concord, Massachusetts. September 2000.

Hartigan, J.P., J.A. Friedman, and E. Southerland, 1983. Post-audit of lake model used for NPS management. *Journal of Environmental Engineering, ASCE*, Vol. 109, No. 6, Pp. 1354-1370. December 1983.

Tracy, J.C., and C.M. Keane, editors, 2000. Peer Review of Models Predicting the Fate and Export of PCBs in the Lower Fox River Below DePere Dam, A Report of the Lower Fox River Fate and Transport of PCBs Peer Review Panel. American Geological Institute, Alexandria, Virginia.

WDNR, 1999. Draft Studies Completed on Cleanup of PCBs in Lower Fox River Sediments. (Published newsletter.) Wisconsin Department of Natural Resources, Madison, Wisconsin. October 1999. http://www.dnr.state.wi.us/org/water/wm/lowerfox/rifs/foxrivercleanup.pdf, accessed April 18, 2001.

Weston, 2000. Supplemental Investigation Work Plan for the Lower Housatonic River. Roy F. Weston, Inc., Manchester, New Hampshire. Prepared for the U.S. Army Corps of Engineers, New England District, Concord, Massachusetts. February 2000.